Models of Opinion Dynamics and Mill-Style Arguments for Opinion Diversity

Baumgaertner, Bert

Historical Social Research
Historische Sozialforschung

Bert Baumgaertner:

Models of Opinion Dynamics and Mill-Style Arguments for Opinion Diversity

doi: 10.12759/hsr.43.2018.1.210-233

Published in:

Historical Social Research 43 (2018) 1

Cite as:

Baumgaertner, Bert. 2018.
Models of Opinion Dynamics and Mill-Style Arguments for Opinion Diversity.
Historical Social Research 43 (1): 210-33. doi: 10.12759/hsr.43.2018.1.210-233.

Historical Social Research
Historische Sozialforschung

All articles published in HSR Special Issue 43 (2018) 1:
Agent-Based Modeling in Social Science, History, and Philosophy.

Models of Opinion Dynamics and Mill-Style Arguments for Opinion Diversity

Bert Baumgaertner *

Abstract: *»Agentenbasierte Modelle der Meinungsdynamik und das Mill'sche Argument für Meinungsvielfalt«.* John Stuart Mill advocated for increased interactions between individuals of dissenting opinions for the reason that it would improve society. Whether Mill and similar arguments that advocate for opinion diversity are valid depends on background assumptions about the psychology and sociality of individuals. The field of opinion dynamics is a burgeoning testing ground for how different combinations of sociological and psychological facts contribute to phenomena that affect opinion diversity, such as polarization. This paper applies some recent results from the opinion dynamics literature to assess the impacts of the Millian suggestion. The goal is to understand how the scope of the validity of Mill-style arguments depends on plausible assumptions that can be formalized using agent-based models, a common modeling approach in opinion dynamics. The most salient insight is that homophily (increased interactions between like-minded individuals) does not sufficiently explain decreased opinion diversity. Hence, decreasing homophily by increasing interactions between individuals of dissenting opinions is not the simple solution that a Millian-style argument may advocate.

Keywords: John Stuart Mill, opinion dynamics, opinion diversity, homophily, polarization.

1. Introduction

I'm going to show how models of opinion dynamics allow us to better understand John Stuart Mill's arguments concerning diversity of opinion. In particular, I will demonstrate how Mill-style arguments about the benefits of opinion diversity depend on relevant assumptions about the psychology and sociality of individuals, thereby restricting the scope of validity of said arguments. Just as logic can be used to formally investigate and rigorously analyze arguments, so simulations can be used to explicitly codify assumptions underlying arguments, in this case the benefits of opinion diversity.

* Bert Baumgaertner, Department of Politics and Philosophy, University of Idaho, 875 Perimeter Drive, MS 3165 Moscow, USA; bbaum@uidaho.edu.

The growing field of *opinion dynamics* studies how opinions change over time as a function of psychological and sociological assumptions. Researchers in opinion dynamics have been primarily interested in how consensus or polarization at the population level emerge from interactions at the individual level, but their insights also apply to diversity of opinion. For example, the maintenance or generation of diversity of opinion can be seen as the failure of reaching consensus or polarization: in many cases modelers start with a population of randomly distributed opinions (which is presumed to capture a certain sense of what is meant by diversity – discussed later) and then see how that distribution changes as interactions take place in the model.

The field of opinion dynamics is not directly an outgrowth of Mill's work, nor is it housed as a specialization within philosophy. It is nevertheless relevant. If it can be said that there is a specialization in philosophy called 'formal social epistemology,' then opinion dynamics would most closely resemble it. I say 'close' because there is one key difference. Social epistemology (not just the formal investigations) keeps a close eye on how social practices affect the spread of *true* beliefs, while opinion dynamics models tend to abstract truth from its analyses. There are alternative approaches in social epistemology called 'social doxology' that also do not make truth a primary focus of their analyses, and comparisons between veritistic social epistemology and social doxology have been made.[1] However, I am not concerned in this paper with the relationships between these areas. My goal is to draw attention to some of the insights that the opinion dynamics literature provides and show how they bear on philosophical thinking of opinion diversity.

Another closely related area includes a body of work in philosophy of science on the cognitive division of labor (see for example Kitcher 1990; Hegselmann and Krause 2006; Weisberg and Muldoon 2009; Zollman 2010; and the paper by Borg et al. 2018, in this HSR Special Issue). This body of work reflects a growing interest in the social aspects of belief acquisition. However, the emphasis in the research on the division of cognitive labor is different than in opinion dynamics. The former models represent investigations of scientists and their research programs, whereas the opinion dynamics literature tends to represent populations at large. Thus, the decisions for how to model the cognitive aspects of individuals differ substantially. For example, although both areas take bounded rationality seriously, philosophers of science are less interested in studying the effects of cognitive biases or explaining phenomena such as polarization.[2]

[1] See the first part of Goldman (1999), where social doxology is criticized as an alternative to veritistic social epistemology.

[2] There are of course exceptions. For example, it can be shown that polarization emerges even when agents are boundedly rational. See, for example, ongoing work by members of the Computational Social Philosophy Lab at the University of Pennsylvania.

The results from the opinion dynamics literature also have political relevance. Polarization, for example, comes in many flavors, including increased partisan voting patterns, the adoption of more extreme policies, the rise of overtly partisan television networks, or more generally the divergence of political attitudes. It has been argued that political polarization is caused by *homophily*, i.e., increased interaction between like-minded individuals (Gilbert, Bergstrom and Karahalios 2009; Baron et al. 1996; Sunstein 2002). Arguments of this sort appeal to empirical studies to support their claims. However, some findings from the opinion dynamics literature demonstrate that homophily does not produce polarization unless we make additional assumptions about how individuals update their opinions (Dandekar, Goel and Lee 2013). In other words, it is possible to have populations with homophily that fail to be polarized, and hence an appeal to homophily is not a sufficient explanation for polarization. This point will be discussed in more depth later.

Most of the models discussed in this paper are known as agent-based models (ABMs).[3] The main idea is that these models explicitly represent individuals of a population and their interactions with one another. There is sometimes a temptation to think of agent-based models as 'more realistic' than their equation-based counterparts. (Equation-based models represent aggregates of individuals and how those aggregates change through time or space.) No assumption that ABMs are more realistic will be made here. The insights that we glean from agent-based models of opinion dynamics should not be thought of as providing additional realism to the ideas that are being tested. Rather, the application of agent-based modeling in this context makes salient assumptions that Mill and others hadn't considered and shows that these assumptions are relevant to the arguments that are forwarded. These assumptions need not bear on reality directly, but they can guide modeling choices by affecting the behavior of models. Section 5 discusses in more detail how ABMs expose important assumptions without presupposing that they do so by better representing reality.

A brief note on terminology. I will use 'belief' and 'opinion' interchangeably. Contemporary philosophy tends to favor the term 'belief.' John Stuart Mill in his work *On Liberty*, from which I will draw, used the term 'opinion.' The modeling literature that I will draw from also tends to use the term 'opinion.' There is no substantial difference in how these terms are applied in these different areas.

The rest of the paper is organized as follows. In Section 2 I provide a summary of the main concepts that will be relevant to subsequent sections of the

3 Readers may also have heard of individual-based models (IBMs). The difference between them does not concern us here. Besides, the difference seems to be a historical contingency of how the same methodology has been named: 'IBM' is the preferred term in areas like ecology, whereas 'ABM' is more common in areas like economics and other social sciences.

paper. In particular, I provide the background for understanding the role that homophily plays in explaining polarization – contrary to many explanations, homophily is *not* sufficient for producing polarization. In Section 3 I use results from the opinion dynamics literature to argue the point that phenomena such as polarization are generated by the right *combinations* of psychological and social factors. Then in Section 4 I show how these considerations bear on Millian arguments for diversity of opinion. Specifically, I argue that we should view opinion diversity as a kind of process, not as a kind of distribution of opinions. The reason for this, in brief, is that in conditions where the truth (or 'ideal opinion') is obvious, we expect *consensus* of opinion rather than diversity. But when the truth (or 'ideal opinion') requires some type of searching, then the strategies of individuals and the processes that govern opinion updating ought to yield diversity of opinion. These philosophical reflections in turn encourage the development and investigation of opinion dynamics models that incorporate the role of truth. Finally, in Section 5 I take up the topic concerning realism and the agent-based models discussed in this paper.

2. Background

2.1 Millian Arguments

John Stuart Mill advocated that dissenting opinions should not be silenced. By interacting with someone that disagrees with us, one of two things will happen. Either our interlocutor will show that our opinion is mistaken, in which case we trade a false opinion for a true one, or, in order to convince our interlocutor that they are mistaken, we will have to provide reasons for our opinion, and thereby better understand why our opinion is true. In either case, Mill argues, we stand to benefit from seeking out dissenting opinions. This argument can be found in *On Liberty*:

> But the peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error. (Mill, Chapter 2 "Of the Liberty of Thought and Discussion," On Liberty)

Expressed succinctly, one reason that dissenting opinions should not be silenced is that they deprive individuals from either correcting false beliefs (if the dissenting opinion turns out to be true) or better understanding why their beliefs are true.

Two points concerning this argument are relevant here. First, the primary motivation for not silencing the expression of dissenting opinions is centered around a concern for the formation of true beliefs. Second, the argument does not explicitly advocate that individuals actively seek out or disseminate dissenting opinions, but rather that dissenting opinions deserve at least a passive existence. These points, however, are not essential to Mill's thinking about the importance of dissenting opinions. Mill provides a similar argument elsewhere that does not depend on a concern for truth and expresses an active promotion of interactions between dissimilar persons. In *Principles of Political Economy with some of their Applications to Social Philosophy*, Mill says:

> But the economical advantages of commerce are surpassed in importance by those of its effects which are intellectual and moral. *It is hardly possible to overrate the value, in the present low state of human improvement, of placing human beings in contact with persons dissimilar to themselves, and with modes of thought and action unlike those with which they are familiar.* Commerce is now what war once was, the principal source of this contact. (Mill, 1848, Book III, Chapter XVII, Paragraph 1, emphasis mine)

In this argument, Mill makes room for opinions that may not be straightforwardly true or false, but rather express values, morals, or ideals. The existence of dissenting opinions, then, can have pragmatic advantages (e.g. economic) or otherwise. Furthermore, this argument fosters a more proactive approach in that it instructs or encourages us to have interactions with persons that are dissimilar to ourselves; it advocates that we actively seek out dissenting opinions, not just allow for their passive existence.

Across these arguments, the central idea in Mill's thinking is that increasing interactions between individuals with dissenting opinions is somehow advantageous. Mill provides us with some reasons for thinking this, and other Mill-style arguments can be formulated in the same spirit that tout the advantages of opinion diversity. These arguments, however, rely on some mental model of how interactions between individuals (dissenting or otherwise) spread opinions through a population. Not all mental models, even plausible ones, will produce favorable assessments of Mill-style arguments. For example, if interactions between dissenting opinions results with sufficient frequency in the backfire effect, where individuals actually strengthen their initial opinion as a result of the interaction rather than be more disposed to changing it, then opinions across the population will change less over time (or even polarize) and thereby stall whatever advantage the relevant opinions are supposed to bring.[4] So assessment of Mill-style arguments depends on knowing more of the details of the intended model (and whether those details hold true for real populations).

In order to make salient the relevance of some details, we can start by outlining some features of a model explicitly. Let m be some kind of metric by which populations can improve or diminish with respect to, be that an increase

[4] The first discussion of the backfire effect (by that name) appears in a political context in the work of Nyhan and Reifler (2010).

in wealth, the frequency of true beliefs, or otherwise. Let *r* be a source that shapes people's opinions independent of social influence. For example, *r* might be a reason that individuals generate for themselves, contemplate, and then change their opinion in response to. We can think of *r* like a resource by which opinions are changed in such a way that *m* is increased. According to Mill-style arguments, some patterns of interactions will be better at distributing *r* through the population than other types of interactions. Specifically, populations where interactions occur primarily between like-minded individuals will be less adept at distributing *r* than populations with increased interactions between individuals of dissenting opinions. This could be because the interaction between dissenting opinions is a way of generating *r* ("the clearer perception and livelier impression of truth, produced by its collision with error"), or because such interaction is a way of extending or spreading *r* to individuals that did not previously share the opinion that helps increase *m* ("the opportunity of exchanging error for truth").

A sketch of this model should allow for opinions to change even in the absence of *r*. That is, while *r* may be one source of opinion change, there are other processes by which opinions will spread, such as social influence. For example, absent *r*, the changing of opinions may drift with respect to *m*, sometimes increasing it and sometimes decreasing it. Other processes of opinion change, however, may cause rampant polarization, preventing increases in *m* even with the introduction of *r*. Mill-style suggestions can be understood as an attempt to describe how these processes of opinion change, which can be investigated separately from *r*, come to affect the impacts of *r* when *r* is present. That is, we can ask how processes of opinion change can be conducive in getting *r* to increase *m* by spreading *r* through the population. For example, if polarization is contrasted with opinion diversity, then we can understand Millian suggestions as saying that the processes by which opinions change that make a population more adept to the spread of *r* are also processes that tend to avoid polarization and have more, e.g., "drift-like" characteristics. This is where the relevance of opinion dynamics lies: in specifying more explicitly these processes of interest in order to better understand their contributions to the spread of opinions.

In turn, we obtain a better understanding of how to assess Mill-style arguments by specifying under what conditions interactions between dissenting opinions is sufficient for avoiding polarization and increasing opinion diversity. Although not stated explicitly, the underlying assumption seems to be that a lack of opinion diversity is caused by not having sufficient number of interactions between non-like-minded individuals. As explained in more detail below, *homophily* occurs when there are far more interactions between like-minded individuals than non-like-minded individuals, and techniques have been developed to measure this property. Put in these terms then, Mill-style arguments suggest that homophily somehow prevents opinion diversity; reversing homophily is a sufficient strategy for improving opinion diversity. The assessment of whether Mill is correct will turn out to be "it depends" and we will be in a better position to say on what.

2.2 Psychological and Sociological Background

The processes of opinion change and spread can be divided into the psychological and the sociological. The Millian suggestion is directed towards sociological changes – it does not, for example, promote a change in *how* opinions or beliefs should be updated. With respect to the latter, philosophers are fond of the Bayesian framework, an appropriate idealization given the aim of studying what ideally rational agents would do. Moreover, such an aim gives some license to freeing oneself from the constraints of human psychology. However, since the Millian strategy is geared towards making certain types of *interactions* more frequent, i.e., between individuals with differing opinions, and not how someone should update their opinions given an interaction, it is more appropriate to select idealizations about opinion updating that are motivated by what is empirically known about human cognition.

To this end, the relevant literature includes work done on a family of related biases, including biased assimilation,[5] myside bias, and confirmation bias.[6] A common thread of all of these biases is that an individual's current opinion or belief can skew how new information is integrated. For example, information that is consistent with one's opinions is more likely to be adopted, while information that is at odds with one's opinion is discounted, or even ignored. For the points that will be important later, the differences between these biases will not matter so much as the family of them.

The sociological processes are about how interactions are "selected." Interactions throughout a population are structured by factors related to the mediums by which opinions are made accessible. For example, how opinions spread through a society that trades information strictly through face-to-face interactions is going to be different than a society that has access to the internet. For one, the internet removes physical limitations that may otherwise prevent individuals from interacting, which suggests at first that the number of interactions or connections between individuals in the population increases. That is, the introduction of the internet generates the expectation that the number of enclaves or echo chambers would drop as a result of removing physical limitations of communication. However, contrary to this expectation, a society with access to the internet is not simply an increase in scale of connections of communication. While the advent of the internet has increased accessibility to

---

5 In particular, see Lord, Ross and Lepper (1979); Miller et al. (1993); Munro et al. (2002); Taber and Lodge (2006).

6 Jonas et al. (2001); Wason (1960, 1968).

information generally, information is not well-mixed. The internet has significant factions and enclaves where individuals get their information from, which can form, e.g., echo chambers. Opinions, and information generally, are not evenly distributed among real populations and are not equally accessible to any individual.[7]

We can be more precise about the accessibility of information and opinions and how they are distributed. An important advancement in the social sciences and relevant humanities is the recognition of network structures and their effects on how something like information spreads or diffuses in a population. The next section provides a brief introduction to network theory.

2.3 Some Technical Background on Network

A network consists of nodes (or vertices) and links (or edges). Nodes might represent a Facebook or Twitter account, while links could represent friendships or followers. The *structure* of a network refers to the distribution of links among nodes. Numerous distributions of links are possible given a set of nodes, but several statistical patterns tend to emerge for different networks.

There are four common types of networks: random, regular, small-world, and scale-free. In a random network, some number of links are randomly (uniformly) distributed between nodes. With a sufficient number of links, such networks are good approximations for well-mixed systems – imagine for example a schoolyard full of children playing tag. By contrast, in a regular network every node has the same number of links. A simple example would be a ring of nodes where each node is a member of two links, one to its left and one to its right. Call the nodes connected by links 'neighbors.' A ring can be extended into another regular graph by having nodes form links with their neighbors' neighbors. Another example is wrapping the edges of a chessboard such that columns A and H are connected and rows 1 and 8 are connected (if you're imagining this in three dimensions, it would form a very thick donut shape). A small-world graph is one where most nodes do not have links between them, but only a small number of traverses over links are required to form a path from one node to another. A simple example would be taking a ring of nodes and adding some links between nodes that are "across the pond." Finally, a scale-free network is one where the distribution of links follows a power law, that is, the frequency of nodes that have very few links will be high while the frequency of nodes with many links will be low. Many social networks are scale-free: a few celebrities have lots of Twitter followers, while most people have relatively few.

7 For an accessible introduction to this area that draws from numerous fields, see Hendricks and Hansen (2016).

The networks just described tend to be static, that is, the links between nodes don't change over time. They are thus good approximations for systems where connections between individuals change much more slowly than the diffusion processes that happen on them. In some cases, however, the connections between individuals change frequently enough, and even change in response to what's happening in the system. In these scenarios it is important to model changes in links. For example, people may change who they come in contact with during the course of an epidemic, and then return to their usual interactions once the epidemic dies down. Models of such phenomena are improved by using *adaptive networks* instead of static ones, which has spawned a growing area of study. However, the details here would take us too far from more relevant considerations.

2.4 Homophily

An important pattern found in many networks is *homophily*. Homophily is the tendency for there to be more links between like-minded nodes and fewer links between non-like-minded nodes. Many social networks exhibit homophily.[8] In fact, it has been argued that homophily results in polarization, such as increasing partisan voting patterns, the adoption of more extreme policies, the rise of overtly partisan television networks, or more generally the divergence of political attitudes.[9] Such arguments, while based on empirical studies, wave their hands towards the underlying explanation. They do not attempt to understand the underlying causes or mechanisms of how homophily could produce polarization. As we will see shortly, it is possible to have populations with homophily that fail to be polarized. Additional assumptions need to be made in order to infer that homophily is a cause of or factor in producing polarization. Making explicit these sorts of assumptions is where the field of opinion dynamics plays an important role in the assessment of Mill-style arguments. For example, interactions between individuals with differing opinions can be approximated with models where individuals mix randomly (i.e., where homophily is low). But as we'll also see, the mixing assumptions by themselves are not sufficient to avoid polarization, thus reducing the scope of validity of Millian-style arguments.

8 See McPherson, Smith-Lovin and Cook (2001); Smith, McPherson and Smith-Lovin (2014); Halberstam and Knight (2016).
9 See Gilbert, Bergstrom and Karahalios (2009); Baron et al. (1996); Sunstein (2002).

3. Generative Opinion Dynamics

In general, an important aim of the field of opinion dynamics is to formally represent how opinions form and spread through real populations. The field of opinion dynamics is growing steadily, but not uniformly in a given discipline. The most technical and formal investigations tend to happen in areas such as physics, where the focus often lies on representing population-level patterns and finding conserved quantities.[10] We could call this sub-field *phenomenal* opinion dynamics. Other areas focus more on how certain population-level phenomena, such as polarization, consensus, or opinion diversity, *emerge* or are *generated* from micro- or individual-level processes. Ideally, these processes are empirically motivated from research in fields such as social psychology. This subfield could be called *mechanistic* or *generative* opinion dynamics.[11] I do not believe there is a bright line that distinguishes generative and phenomenal opinion dynamics, it is rather a distinction that helps highlight the agent-based approaches against a much longer tradition of equation-based modeling, a point I return to later in Section 5.

One feature of generative opinion dynamics is the explicit and separate representation of the psychological and social processes. Representing the psychological and social aspects separately makes it easier to understand their individual contributions and their combined effects. This understanding makes it possible to better assess arguments about the causes of phenomena such as polarization.

For example, as mentioned above, homophily – increased interactions between like-minded individuals – has been cited as the cause for polarization, particularly with respect to partisan issues. Pointing at homophily as an explanation for polarization, however, passes over background assumptions that may be crucial to the explanation. Models allow us to test how such a putative explanation is sensitive to those assumptions. For example, for what range of opinion-updating processes will homophily produce polarization? Will polari-

10 For an overview, see Castellano, Fortunato and Loreto (2009).

11 I shall prefer the term 'generative' over 'mechanistic.' While 'mechanistic' has the right flavor in that it draws attention to the mechanisms and processes that constitute or cause a phenomenon, specifically to an emergent phenomenon, I want to avoid the reader from thinking that the relevant explanations are mechanistic explanations in the way that the New Mechanists have characterized them (see Bechtel 2008; Craver 2007; Craver and Darden 2013; Machamer, Darden and Craver 2000). It may be possible to extend the standard mechanistic conception of explanation to include some of the complexity that fails to be captured by the early mechanistic conception (see Roe and Baumgaertner 2016). Even if it could, the broader framework of generative explanations is better fitting. In brief, the idea is that if you didn't grow the emergent or macroscopic phenomenon, then you didn't explain it. Generative explanations are described in the work of Joshua Epstein (see e.g., Epstein 2006).

zation occur if people update their opinions by taking some kind of average between theirs and their interlocutors'?

Dandekar, Goel and Lee (2013) show that homophily is not sufficient to produce polarization. That is, homophily will not always produce polarization, it matters what the opinion formation process is, i.e., how agents integrate information. If the opinion formation process is similar to a DeGroot-like averaging process (where opinions are updated by taking a weighted average of their opinion and that of their neighbors), then Dandekar, Goel and Lee demonstrate that polarization is never produced.[12] In order to produce polarization, homophily needs to be combined with something like *biased assimilation*, where individuals discount dissenting opinions and weight more heavily opinions consistent with their own.[13]

A similar lesson can be found in Baumgaertner, Tyson and Krone (2016). They study a relatively weak form of bias they call *amplification* – opinions can become more entrenched when shared between like-minded individuals.[14] They introduce an attitude spectrum, {-L, ...,-2, -1,1,2,...,L}, where the sign of the attitude represents a yes/no opinion and the magnitude of the attitude represents how entrenched that opinion is. By default, agents update their opinion by increasing or decreasing their attitude towards their interlocutor, which over several iterations would make agents "meet in the middle."[15] But, with some small frequency, agents will "amplify" their opinions when interacting with an agent with the same opinion (how frequently this occurs can be varied in the model). When the opinion formation process without amplification is implemented on a regular network (e.g. a grid), opinions will tend to form clusters that become larger over time. The distribution of attitudes, however, moves towards centered opinions. That is, extreme opinions disappear over time until only the -1 and 1 attitudes remain (and then eventually, after a very long time, a single attitude will remain, though which one will be subject to chance). The clustering of opinions is something that emerges in their model and is a representation of homophily. Such clustering, however, is not enough to produce

12 "DeGroot-like" refers to the work of DeGroot (1974).

13 To be clear, the claim is not that it is impossible to produce polarization by having agents completely cut off links to dissenters and only create links to like-minded people; that is in principle possible. Rather, the claim is that empirically-informed levels of homophily will still contain a sufficient number of links between non-like-minded individuals such that polarization is not achieved if individuals update their opinions by some DeGroot-like averaging process.

14 Amplification is a stylized version of a bias related to a family of biases where one's initial opinion or belief biases subsequent opinion updates. Examples include biased assimilation, the myside bias, or confirmation bias. Studies on biased assimilation include Lord, Ross and Lepper (1979); Miller et al. (1993); Munro et al. (2002); Taber and Lodge (2006). For empirical studies on confirmation bias see Jonas et al. (2001); Wason (1960, 1968).

15 This strategy closely resembles the conciliationist position in the peer disagreement literature.

polarization. In their case, generating polarization occurs with (in combination with the emergence of clustering) the addition of a very small amount of amplification.

Given the results of Baumgaertner, Tyson and Krone, and Dandekar, Goel and Lee, the temptation may be to point towards something like amplification or biased assimilation as the root or dominant cause of polarization. Some opinion models have explored this possibility. Bounded confidence models are one example that show how psychological features alone could produce polarization.[16] These models are "well-mixed" systems in that interactions are unstructured – agents interact with each other randomly. That does not mean, however, that all agents can influence each other. The notable feature of bounded confidence models is that agents that become sufficiently dissimilar with respect to their opinions cease to affect one another. In bounded confidence models, agents have continuous opinions (from -1 to 1) and whether one agent changes her opinion in response to another's depends on whether the distance between their opinions is below some threshold. This threshold is interpreted as something like uncertainty or bounded confidence around the opinion. If the uncertainty of one opinion overlaps with the uncertainty of another, then the agents update their opinions, otherwise no updating occurs.

In variations of the bounded confidence model, sometimes referred to as relative agreement models, an agent changes her opinion linearly in response to another with the amount of overlap between the opinion segments (the amount of their agreement) as well as their confidence. For example, the opinion of an agent with high confidence (i.e., low uncertainty) will occupy a narrow segment of the opinion spectrum, while an opinion of an agent with low confidence (high uncertainty) will occupy a larger segment of the opinion spectrum. Thus, two confident agents are less likely to have their opinion segments overlap while agents with high uncertainty are more likely to overlap in their opinion segments. Moreover, the uncertainties of opinions decrease as agents interact with others whom they agree with. In a sense, agents can become more "close-minded" over time because they do not "listen" to opinions that are sufficiently different.

One of the interesting results of bounded confidence models is that a population of agents with initially identical uncertainties but randomly distributed opinions can nevertheless converge towards a fragmented distribution of opinion – they fail to converge on one opinion. For example, if the initial uncertainties are not too high, then a population of agents with opinions randomly distributed between -1 and 1 can converge towards two clusters, one cluster around -0.5 and the other around 0.5. Deffuant et al. (2002) show how populations with extreme opinions – opinions that are held with sufficiently high

[16] Notable papers include Hegselmann and Krause (2002); Deffuant et al. (2002); Dandekar, Goel and Lee (2013); Salzarulo (2006); Deffuant et al. (2000); Weisbuch et al. (2003).

confidence towards the polar ends of the spectrum (- 1 and 1) – can produce something like polarization. That is, agents that are more moderate in their opinion (around 0) and have sufficiently high uncertainty, will eventually be drawn towards one of the polar ends of the spectrum. While these results hold for well-mixed populations, it is expected that the results also hold (if not exacerbated) if the population were on a network, i.e., if the interactions were structured.

The results discussed so far may tempt us to think that polarization is the product of some set of psychological features of opinion updating, that we can (at least for the most part) ignore the social aspects of opinion dynamics. It is not always the case, however, that when psychological features produce population-level phenomena in the case of well-mixed populations, that the same results hold for structured interactions. For example, in their follow-up work, Baumgaertner, Tyson and Krone (2017) demonstrate that if populations are sufficiently mixed, then amplification actually *decreases* the time it takes to reach consensus – polarization is avoided entirely. By "mixing" they have in mind that agents have both interactions with local neighbors, but also some interactions with random agents in the population (this is in effect a small-world network). Thus the same psychological processes produce wildly different results depending on the structure of interactions.

So, what is the lesson we are to take so far? Opinion dynamics models help us recognize that certain population-level phenomena like polarization emerge only with the right *combination* of population structure and psychological features of opinion updating. Homophily by itself doesn't produce polarization, you need to add biased assimilation or amplification. But biased assimilation or amplification by themselves won't necessarily produce polarization either. Interactions have to be sufficiently structured in a certain homophilic way in order for a family of biases to produce polarization. Then again, some opinion formation processes, like those described by bounded confidence models, will produce polarization for both well-mixed and structured populations. But these assume that individuals become increasingly "closed-minded" over time. The point is that it matters what we are assuming with respect to social and psychological aspects of beliefs when we are imagining how they spread and whether the end result will be consensus, polarization, or diversity. What all this means for Mill-style arguments about diversity is treated in the next section.

4. Lessons from Opinion Dynamics Models

The previous section contrasted sociological aspects of opinion formation against psychological ones and provided some examples where different combinations of these features affect phenomena such as polarization and consensus differently. We are nearly in a position to say how these considerations bear on the study of opinion diversity and its importance to Mill-style arguments. But before we do, we need some additional clarity on what is meant by 'diversity' in this context.

4.1 Distribution vs. Process View of Diversity

Is diversity some distribution of opinions or some structured relationship between them? Or is it better to think of diversity in terms of processes? Consider the case of polarization. On the one hand, polarization can refer to a distribution of opinions with higher frequencies on the ends of a spectrum and low frequencies in the center. Call this the distribution view. On the other hand, polarization can refer to the processes that would, under suitable conditions, produce clusters of opinions on the far ends of an opinion spectrum. Call this the process view. There is no standard interpretation of polarization in terms of either the distribution view or the process view. The same difference can be applied to diversity. I will argue that we should think of diversity in terms of processes and not distributions.

The point of arguing for either a distribution or process view matters to our understanding of Mill-style arguments for opinion diversity. If we adopt the distribution view of diversity, we find ourselves in a (near) paradoxical position, while the process view avoids this. Consider what it means to produce or maintain opinion diversity on the distribution view. On the one hand, we want diversity of opinion in order to maximize our chances of finding the truth (or whatever the relevant measure *m* is, as discussed in Section 2.1 – for simplicity's sake I will use truth as the illustrative case). On the other hand, we expect that epistemic peers would reach consensus, and thus see a decrease in diversity. On the distribution view then, diversity of opinion is not the thing we should be aiming for because it would encourage false beliefs. If we are after the truth, we want populations to be receptive and adaptive to information. Having a diverse set of opinions initially is a good way to start, but the maintenance of diversity can come at the cost of our ultimate aim: consensus on the truth.

The process view of diversity avoids this issue. It matters less what the distribution of opinions looks like than what the distribution would look like under suitable conditions. If it is not obvious what the truth is, or if we are in a domain where truth is not the relevant notion, as may be the case when it comes to matters of mere taste, then diversity of opinion is expected to create a distribution of opinions that is relatively uniform on an opinion spectrum. If, however, the truth is as bright as the light of day, then we expect opinion diversity to be overshadowed by the truth. The interesting point of consideration lies between these two scenarios, where there are truths about a domain, but they require work to uncover; where truths require some kind of searching. In some cases we may get lucky by starting somewhere where it is easy to stumble in the right area. The supposed benefit of opinion diversity is that it needn't depend on such luck. Even if we happen to start our search in a place that's away from the truth, by moving around in various directions some members of the population may find the truth before others, but ultimately others will find their way to the same place (e.g., through some form of communication).

A second argument for the process view addresses the accessibility of opinions and information. It is possible for a social network with homophily to have a global distribution of opinions that is diverse. Partition an opinion spectrum however you like, but make sure that each segment is represented by agents in the population. Then, set up a network so that links between agents exist when they have the same opinion.[17] In such a scenario the population has a diversity of opinions, but with little to no contact between those differences of opinion. On the process view, such a scenario would not count as diverse (but would on a distribution view). If agents with differences of opinion do not communicate with one another, then it is significantly less likely that any of them will change their mind since they do not come in contact with dissenting opinions.

This is one way to understand the point of one of Mill's argument. His argument describes the importance of seeking out dissenting opinions. We can understand this as an attempt to break down homophily and create links between individuals that are not like-minded. If opinions are distributed in such a way that homophily is decreased, then we may expect to see more opinions to change in what would be presumably a beneficial way. But before we consider the caveats of this idea in more detail, a brief concession should be made for the distribution view.

Although I have argued against the distribution view of opinion diversity, it does have a limited explanatory role. For example, Duggins (2017) argues that one of the shortcomings of the fields of opinion dynamics thus far is its failure to recognize diversity as a population-level phenomenon in its own right. Most of the literature has focused on polarization or consensus, but there are many scenarios where we see a diversity of opinions. If polarization and consensus merit explanations, then so too does diversity. In order to account for this phenomenon, Duggins (2017) introduces the ISC (influence, susceptibility, and conformity) model. In the ISC model, diversity of opinions can be maintained because individuals end up being pulled towards the center and the extremes simultaneously in a population that balances heterogeneous intolerance, susceptibility, and conformity.[18] So one role of thinking of opinion diversity in terms of distributions is to help identify relevant phenomena and then build

[17] To make sure that the network is connected, one can add links required between agents that are different, but only so many as to make the network connected, no more.

[18] The details of these concepts are not relevant for the point being made here. Readers interested in better understanding how these notions are formalized in the ISC models are referred to the original article.

models of it. This role, however, is transient. As Duggins goes on to argue using his ISC model, the phenomenon of diversity can be produced by individuals being simultaneously pulled in multiple directions. This is an appeal to processes. Duggins does not consider how these processes might change if some opinions were the true ones. However, if truth were incorporated in his model somehow, presumably we expect that diversity of opinions will be temporary and that at least in the limit the population can converge on the truth.

4.2 On Millian Arguments for Diversity

Let us finally turn to some insights we glean from our considerations of opinion dynamics models as they apply to Mill-style arguments for opinion diversity. To briefly recap, the central idea of Mill is that increased interactions between people of different opinions is a sufficient strategy for increasing some metric by which a population improves as a result of the relevant opinion spreading (e.g., more true beliefs). What our consideration of opinion models provide are numerous caveats for, and refinements of, this idea, particularly in those scenarios where it is not obvious what, e.g., the truth is, and the spread of opinions is largely governed by, e.g., social influence.

Let us suppose that we have a population that already has a diverse range of opinions. Under what conditions can that diversity (or something close to it) be maintained as agents update their opinions, particularly when a small bias exists in which agents tend to amplify their opinions? Baumgaertner, Tyson and Krone (2016) provide one answer: in the fully spatial case (where the network is regular), diversity of opinion or attitudes can be maintained by counterbalancing amplification with co-influence (or "centrist") functions. Centrist influence functions in their model are a way to bias interactions with agents that hold moderate or center opinions. Thus, while amplification and clustering tend to cause agents to become more extreme in their opinions and thereby cause polarization (not diversity), centrist influence functions pull agents towards the center of the spectrum. In this sense their model is similar to the model developed by Duggins: diversity of opinion is maintained when agents are pulled in two directions. The important point to notice here is that it is not simply having any non-like-minded individuals interact that ensures diversity, but specifically increased interactions between more extreme opinions and moderate opinions. This idea is consistent with the mental model behind Millian-style arguments, but provides refinement of those ideas.

What if, instead, a population is initially polarized? What processes lead to the undoing of polarization and towards a more diverse distribution of opinion? An artificial way of doing this would be to "inject" new opinions into the population through the birth of new individuals. New opinions, however, need not come to be in this way. If we imagine that the population remains the same through time, what needs to be avoided is the possibility that sufficiently many agents can become too close-minded. That is, it should not be the case that sufficiently many agents reach a point where they no longer change their mind. As we saw in the case of Deffuant et al. (2002), close-mindedness, i.e. extremely high confidence in an opinion, prevents individuals from updating their opinion unless the other opinion is sufficiently similar, thereby causing and maintaining a polarized population. In this situation, even if the agents interact with non-like-minded individuals, polarization is not reversed. That's because homophily is not the sufficient reason for the polarization, but rather the confidence levels that individuals have achieved after numerous opinion updates that prevent them from "listening" to others. So in this case, the Millian idea to increase interactions between people of different opinions will not increase diversity. We thus have an example where Mill's argument is not valid under one set of plausible psychological assumptions.

In other models (e.g., Baumgaertner, Tyson and Krone 2017) increasing the overall mixing of individuals increases the number of interactions between non-like-minded individuals and can be a means of breaking down polarization; without a sufficient amount of mixing, interactions are structured (i.e., individuals interact with the same set of neighbors) and over time this generates homophily and polarization. However, if the amount of mixing is sufficiently high, another effect takes hold: consensus on a random opinion! If we are thinking in terms of the process view of opinion diversity, then consensus too is at odds with the Millian suggestion. The moral that Baumgaertner, Tyson and Krone (2017) draw is that there is a "Goldilocks" level of mixing: too little and the population polarizes, too much and the population reaches consensus, but just the right amount of interactions between non-like-minded individuals generates and maintains diversity. So there can be too much of good thing – it is not always beneficial to keep increasing interactions between non-like-minded individuals. Here again we see a narrowing of the scope of validity of Mill's argument. Mill's suggestion is correct that, in a scenario where mixing is very low to non-existent, increased interactions help achieve opinion diversity. But in scenarios where there already is some mixing, the suggestion breaks down.[19]

As a final point in this section, allow me to turn the table and demonstrate how philosophical considerations can feed back on opinion dynamics modeling. An opinion dynamics model that exhibits opinion diversity would take clustered populations of opinions and increase their range, but would also make sure that, upon discovering better ideas or the truth, make the variation decrease again – this is an insight we gain from thinking of diversity as a process

---

19 While it can be said more precisely what these levels of mixing are in the model, it is difficult to assess how these levels correspond to real populations. This is one of the challenges of validating opinion dynamics models and requires sufficiently robust data that is difficult to obtain.

not as a distribution. A helpful analogy is ant foraging. When no food has yet been found, ants move around randomly searching for food. When some ants find food, they head towards their nest, leaving behind a pheromone trail. Other ants that encounter the pheromone trail follow it, heading in the opposite direction of the nest and thereby increasing their chances of finding the food source. Once they find the food source, they too return to the nest and reinforce the pheromone trail. Eventually a large fraction of the ant colony can be found walking along the trail, either towards the food, or towards the nest, creating an ant highway. When the food source runs out, however, the ants stop heading back to the nest. The failure of reinforcing the pheromone trail means that the trail eventually dissipates and the ants go back to a random search.

Obviously, humans are not merely ants in the way they form their opinions and beliefs. The point of the analogy is that the population level patterns in the ant foraging case have some important similarities to how we expect a population of humans will react to truth or improved ideas: as the strength or salience of the truth increases, for example, we expect a larger fraction of the population to be drawn towards it, but if that trail is or becomes non-existent, the population ought to return to a wider or more diverse search area (as per the Millian suggestion). To my knowledge, no such opinion dynamics models currently exist and this is an opportunity for new areas of investigation.

Towards a concrete suggestion for the opinion dynamics literature, recall some of the model features outlined in the discussion on Mill, specifically that r ("the clearer perception and livelier impression of truth") is like a resource that changes opinions in such a way that m is increased (some metric by which populations can improve). When r is absent, the processes underlying the opinion dynamics should cultivate opinion diversity. That way, when r is present, the processes increase the chance and rate that m increases because the opinions of the individuals change accordingly. If r is "strong" enough – think of abundance as being akin to how obvious the value or truth of an idea is – then we expect the opinion processes to play a weaker role in driving how opinions change. The opinion dynamics literature explores the case where r is absent. Extending these models to include r, m, and the requisite assumptions is a natural direction for the field to take.

5. Agent-Based and Equation-Based Modeling

One of the main points of this paper has been to examine some assumptions that are left implicit when it comes to arguing for the importance of opinion diversity. Generative opinion dynamics models, often in the form of agent-based models, provide a suitable framework for investigating these assumptions. It may be tempting to think that generative opinion dynamics models, and perhaps agent-based models more generally, provide a better understanding of phenomena such as polarization because they increase realism. In the final section of this paper I will argue that this is not the proper way to understand the contributions that opinion dynamics models provide. Increasing the number of assumptions does not imply an increase in realism, although a model with more assumptions does imply that the model is more specific (or less abstract). A brief discussion of the nature of models helps make this point clear.

Models are sometimes seen as surrogates for a target system. From this perspective, manipulating the model is like manipulating the target system, but without actually affecting the target system. One reason we may not want to manipulate the target system is concerned with ethics. Biomedical researchers, for example, use mice as models for humans in preclinical trials in order to gather information about the toxicity and efficacy of a new drug – using humans to measure toxicity is not generally considered ethical. Another reason may be that manipulating the target system is not feasible, for physical or political reasons. We cannot, for example, remove so much carbon from the atmosphere in a year to bring the total ppm back to 350 to observe the effects on climate. For that, we use models of the climate.

Agent-based modeling can invite the use of a similar perspective where models are target-system surrogates. This temptation seems to come about when agent-based modeling is contrasted with equation-based modeling. An equation-based model tends to represent features of a target system in terms of aggregates. For example, the Lotka-Volterra equations represent the dynamics of two species that interact, where one species is the predator and the other the prey:

$$\frac{dx}{dt} = \beta x - \kappa xy$$
$$\frac{dy}{dt} = \varepsilon xy - \delta y$$

The variables x and y represent the number of prey and predators, respectively. The β parameter represents the birth or growth rate of the prey (absent predators) and κ represents the rate at which the prey are killed, which is assumed to be proportional to the rate at which predators and prey meet (xy). The ε parameter represents the rate at which the predator population grows from eating the prey and δ represents the natural death rate of the predator population. Notice that all of these terms are representing population-level features of the target system, no individuals are explicitly represented.

An agent-based version of a predator-prey system consists of representations at the individual level. For example, one set of representations, the agents, correspond to the individuals in the target system, the predators and prey. These agents then "interact" by, for example, moving randomly through a two-dimensional space until the coordinates of two agents are sufficiently close (which is determined by some parameter). If the two agents are a predator-prey pair, then there is some probability that the prey agent is "consumed," i.e., the prey agent is removed or deleted, and the "energy" level of the predator is increased. When the energy level of a predator is sufficiently high, it "reproduces," i.e., a new predator agent is brought into existence; when the energy level is sufficiently low, a predator "dies," i.e., is removed or deleted. Finally, a prey agent "reproduces" periodically, bringing a new prey agent into existence.

Several points of comparison between the equation-based and agent-based models are worth highlighting. While the agent-based model has been described in a simple way, anyone who wishes to implement it computationally is faced with numerous choices that must be made. Once made, tests can be done to check whether and how they affect the model (these tests can fall under both sensitivity and robustness analyses). For example, when new predators are brought into existence, what coordinates should they be given? Should they be near the "parent"? Or should new predators be placed randomly? Does this choice make a difference? And with respect to what does it or does it not make a difference? It is unlikely that choosing where to place new predators makes a difference to certain patterns in the dynamics of the two populations, but that would be the case against the backdrop of agents moving around the space randomly. Choices about placement of new predators could very well make a difference if agents aggregated in some way. Similar considerations apply to the introduction of new prey. Of course, some of these details can be avoided by not situating the agents in space at all. Instead, one might have the two populations interact by randomly selecting pairs of individuals. Even still, however, assumptions are being made about the distributions of frequencies of interactions between agents. The point, in short, is that there are many details that have been left out in the short description of the agent-based version that would need to be specified when implementing it.

From the perspective that models are surrogates for target systems, the contrast from equation-based to agent-based models invites the idea that agent-based models are more realistic versions of the target system. This is because agent-based models require more decisions to be made about what details or "parts" to include, where these "parts" can "latch onto the world." It is true that agent-based models make it more difficult to hand-wave over details because the algorithms have to be sufficiently specified in order to implement and run simulations. It would be a mistake, however, to think that this required level of detail implies there are specific things in the world that the "parts" in the model are supposed to latch on to. Take, for example, population structure. In the equation-based model, the assumption is that there is no population structure, the individuals are being "well-mixed." However, there is a sense in which there is an underlying structure, it's just that it is random or stochastic. In an agent-based model, a random network structure (as described above in Section 2) would provide a decent approximation of the well-mixing assumption in the equation-based model. The agent-based model then lets us investigate whether the well-mixing assumption matters to the phenomenon in question. That is, does the model still behave the same way if we assumed instead that individuals interacted in some other structured way? As we saw in the discussion of homophily and amplification, the answer is sometimes "yes." These structures, however, need not bear on reality directly. In many cases they are additional idealizations made to investigate other assumptions or model behaviors. And while network structures have analogs in the world, as discussed in Section 2, we should not be so quick to think that these structures can serve as surrogates for real populations.

Instead, I suggest that we should think of these additional assumptions as characterizing the modeling choices we can expect we would need to make *if* we were to try to create realistic models, models that could be said to be predictive of the target system. Agent-based models can demonstrate the pertinence of several assumptions that are (implicitly) assumed not to make a significant difference in the model behavior. In developing a predictive model, those differences can turn out to matter, and it is worth knowing which choices we have to make and which choices we don't have to make (weighted by other considerations, such as the level of risk if we get the predictions wrong). In short, when models become more realistic they tend to bring more assumptions to the forefront, but models can bring more assumptions to the forefront without being more realistic. That, I suggest, is the role that current opinion dynamics models have in investigating Mill-style arguments.

6. Conclusion

Models can improve our understanding of how different assumptions about psychology and social structure interact to produce patterns of opinion dynamics. The interest here has been to apply the results of the opinion dynamics literature to the Millian suggestion that increasing interactions between non-like-minded individuals is a way to increase opinion diversity, which in turn improves society at large. Specifically, the goal has been to understand how the scope of the validity of Millian-style arguments depends on plausible assumptions that can be formalized using agent-based models. The most salient insight to be gained is that homophily (increased interactions between like-minded individuals) does not sufficiently explain decreased opinion diversity. Hence, decreasing homophily by increasing interactions between non-like-minded individuals does not assure opinion diversity. What we need in order to assess the validity of Millian-style arguments, and more importantly, to obtain the improvements we ultimately strive for, is to understand how opinion dynamics depend on a combination of psychological and sociological processes. We cannot look towards sociological changes alone if they do not join with psychological processes that produce opinion diversity. Understanding which combinations work and which don't is where the opinion dynamics literature provides guidance.

References

Baron, Robert S., Sieg I. Hoppe, Chuan Feng Kao, Bethany Brunsman, Barbara Linneweh, and Diane Rogers. 1996. Social corroboration and opinion extremity. *Journal of Experimental Social Psychology* 32 (6): 537-60.

Baumgaertner, Bert, Rebecca T. Tyson, and Stephen M. Krone. 2017, April. *Spatial Opinion Dynamics and the Effects of Two Types of Mixing*. ArXiv e-prints, arXiv:1704.05012v2.

Baumgaertner, Bert, Rebecca T. Tyson, and Stephen M. Krone. 2016. Opinion strength influences the spatial dynamics of opinion formation. *The Journal of Mathematical Sociology* 40 (4): 1-12.

Borg, AnneMarie, Daniel Frey, Dunja Šešelja, and Christian Straßer. 2018. Epistemic Effects of Scientific Interaction: Approaching the Question with an Argumentative Agent-Based Model. *Historical Social Research* 43 (1): 285-307. doi: 10.12759/hsr.43.2018.1.285-307.

Bechtel, William. 2008. *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Taylor & Francis.

Castellano, Claudio, Santo Fortunato, and Vittorio Loreto. 2009. Statistical physics of social dynamics. *Reviews of Modern Physics* 81 (2): 591-646.

Craver, Carl F. 2007. *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.

Craver, Carl F., and Lindley Darden. 2013. *In search of mechanisms: Discoveries Across the Life Sciences*. Chicago: University of Chicago Press.

Dandekar, Pranav, Ashish Goel, and David T. Lee. 2013. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences* 110 (15): 5791-6.

Deffuant, Guillaume, Frédéric Amblard, Gérard Weisbuch, and Thierry Faure. 2002. How can extremism prevail? A study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation* 5 (4), <http://jasss.soc.surrey.ac.uk/5/4/1.html> (Accessed January 22, 2018).

Deffuant, Guillaume, David Neau, Frédéric Amblard, and Gérard Weisbuch. 2000. Mixing beliefs among interacting *agents. Advances in Complex Systems* 3 (01n04): 87-98.

DeGroot, Morris H. 1974. Reaching a consensus. *Journal of the American Statistical Association* 69 (345): 118-21.

Duggins, Peter. 2017. A psychologically-motivated model of opinion change with applications to American politics. *Journal of Artificial Societies and Social Simulation* 20 (1): 13.

Epstein, Joshua M. 2006. Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton: Princeton University Press.

Gilbert, Eric, Tony Bergstrom, and Karrie Karahalios. 2009. Blogs are echo chambers: Blogs are echo chambers. In *System Sciences*, 2009. HICSS'09. 42nd Hawaii International Conference on, 1-10. IEEE.

Goldman, Alvin. I. 1999. *Knowledge in a Social World.* New York: Oxford University Press.

Halberstam, Yosh, and Brian Knight. 2016. Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter. *Journal of Public Economics* 143: 73-88.

Hegselmann, Rainer, and Ulrich Krause. 2002. Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation* 5: 1-24.

Hegselmann, Rainer, and Ulrich Krause. 2006. Truth and cognitive division of labor: First steps towards a computer aided social epistemology. *Journal of Artificial Societies and Social Simulation* 9 (3): 10.

Hendricks, Vincent F., and Pelle G. Hansen. 2016. *Infostorms. Why do we 'like'? Explaining individual behavior on the social net.* Springer International Publishing.

Jonas, Eva, Stefan Schulz-Hardt, Dieter Frey, and Norman Thelen. 2001. Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology* 80 (4): 557.

Kitcher, Philip. 1990. The division of cognitive labor. *The Journal of Philosophy* 87 (1): 5-22.

Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37 (11): 2098-109.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67 (1): 1-25.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27 (1): 415-44.

Mill, John Stuart. 1909 [1848]. *Principles of Political Economy with some of their Applications to Social Philosophy*, 7nd ed.. London: Longmans, Green and Co.

Miller, Arthur G., John W. McHoskey, Cynthia M. Bane, and Timothy G. Dowd. 1993. The attitude polarization phenomenon: Role of response measure, attitude extremity, and behavioral consequences of reported attitude change. *Journal of Personality and Social Psychology* 64 (4): 561-74.

Munro, Geoffrey D., Peter H. Ditto, Lisa K. Lockhart, Angela Fagerlin, Mitchell Gready, and Elizabeth Peterson. 2002. Biased assimilation of sociopolitical arguments: Evaluating the 1996 US presidential debate. *Basic and Applied Social Psychology* 24 (1): 15-26.

Nyhan, Brendan, and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior* 32 (2): 303-30.

Roe, Sarah, and Baumgaertner, Bert. 2016. Extended Mechanistic Explanations: Expanding the current mechanistic conception to include more complex biological systems. *Journal for General Philosophy of Science* (Online Oct 21, 2016): 1-18.

Salzarulo, Laurent. 2006. A continuous opinion dynamics model based on the principle of meta-contrast. *Journal of Artificial Societies and Social Simulation* 9 (1): 13.

Smith, Jeffrey A., Miller McPherson, and Lynn Smith-Lovin. 2014. Social distance in the united states: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. *American Sociological Review* 79 (3): 432-56.

Sunstein, Cass. 2002. *Republic.com*. Princeton: Princeton University Press.

Taber, Charles S., and Milton Lodge. 2006. Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* 50 (3): 755-69.

Wason, Peter C. 1960. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology* 12 (3): 129-40.

Wason, Peter. C. 1968. Reasoning about a rule. *The Quarterly Journal of Experimental Psychology* 20 (3): 273-81.

Weisberg, Michael, and Ryan Muldoon. 2009. Epistemic landscapes and the division of cognitive labor. *Philosophy of Science* 76 (2): 225-52.

Weisbuch, Gérard, Guillaume Deffuant, Frédéric Amblard, and Jean-Pierre Nadal. 2003. Interacting agents and continuous opinions dynamics. In *Heterogenous Agents, Interactions and Economic Performance,* 225-42. Berlin: Springer.

Zollman, Kevin J. S. 2010. The epistemic benefit of transient diversity. *Erkenntnis* 72 (1): 17.